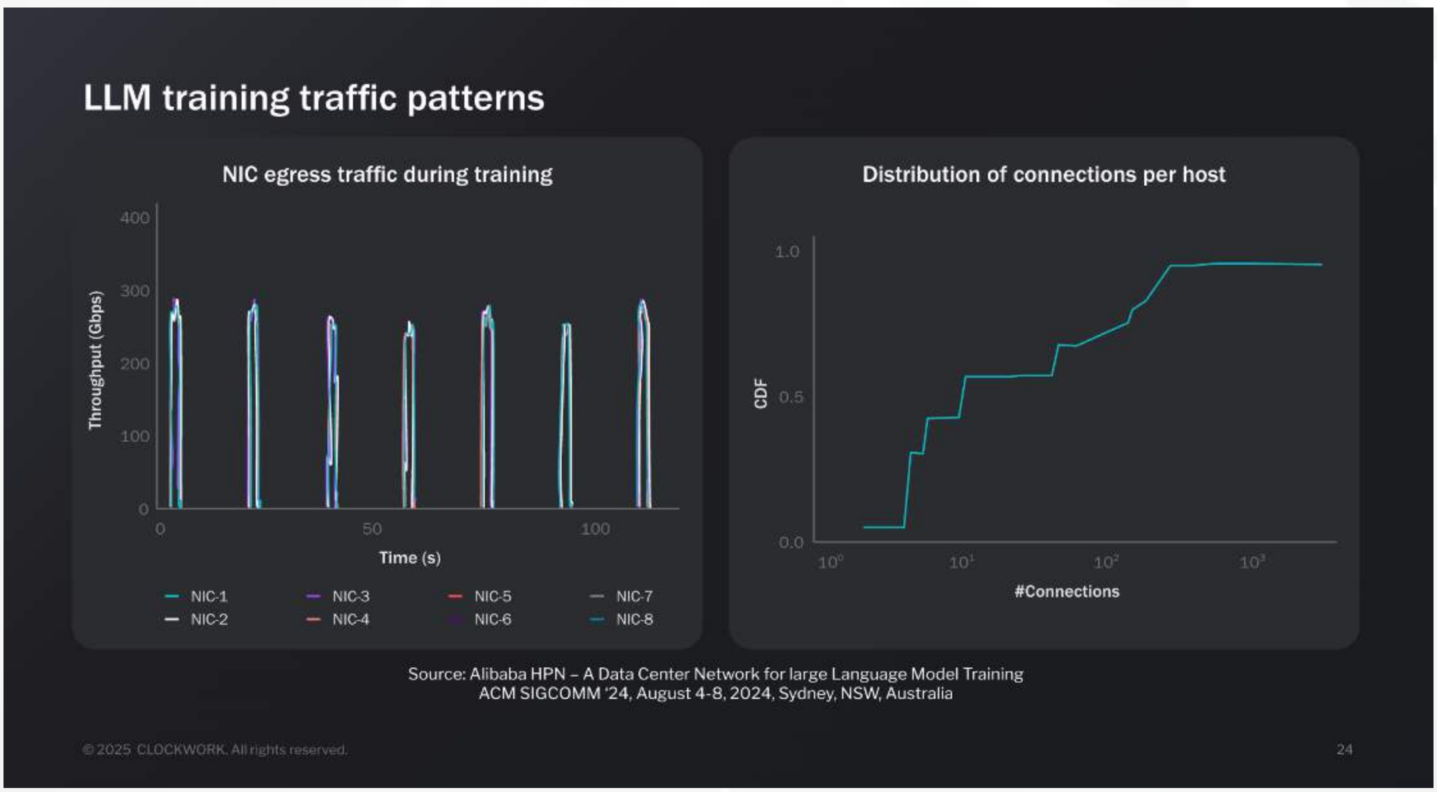
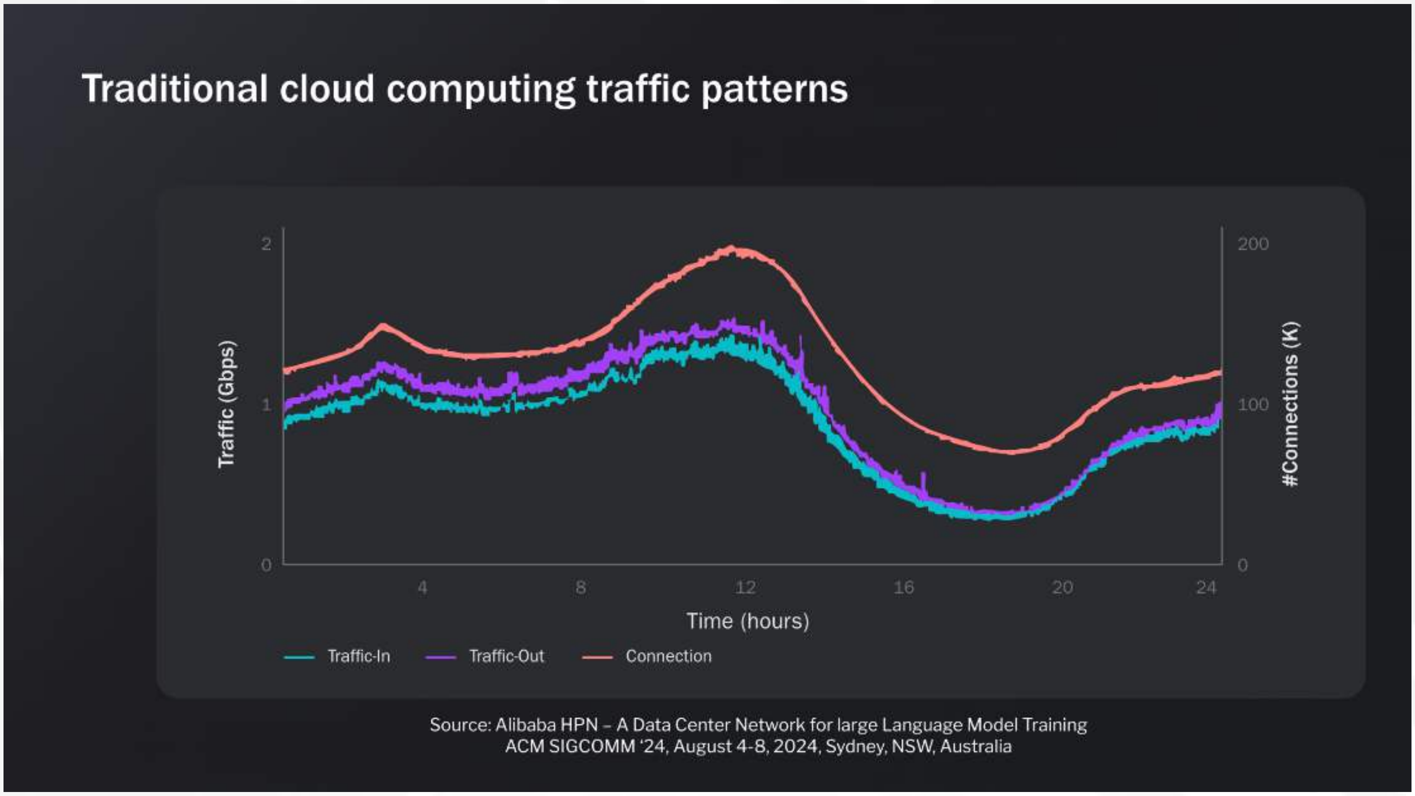
CLOCKWORK.IO

Software-Driven Fabrics

Artificial intelligence (AI) is rapidly moving from specialized research to mainstream applications reshaping entire industries. However, the highly integrated, single-vendor systems that initially powered these advances are now becoming obstacles, limiting innovation and efficiency. History shows that modular, open ecosystems inevitably replace monolithic designs, and the same unbundling is beginning inside AI infrastructure: diverse GPUs, purpose-built accelerators, and Ethernet-based fabrics are displacing single-vendor solutions. Against this backdrop, **Software-Driven Fabric (SDF)** emerges as the networking counterpart to this tectonic shift. SDF enables programmable, vendor-neutral networks designed to manage and optimize data flow dynamically in real-time, particularly for large-scale AI workloads. This paper sets out why SDF is the logical and necessary foundation for the era of heterogeneous, at-scale AI.

AI Workloads Impose Unique Requirements On The Fabric





The rapid increase in AI training and inference over the past few years has resulted in a major upheaval in networking. Modern AI clusters integrate thousands of GPU servers into a single, tightly coupled computer, creating unique network demands that fall into two distinct planes:

- Front-end user facing network:
  - Handles microservices, data ingestion, model outputs, API interactions
  - Latency: milliseconds, moderate packet loss tolerated
  - Dominant protocols: TCP/HTTP, gRPC
- Back-end GPU Network:
  - Supports collective GPU operations (e.g., all-reduce, broadcast)
  - Latency: microseconds, near-zero packet loss required
  - Dominant protocols: RDMA, specialized collective libraries

The ***back-end GPU network*** **is a high-performance fabric dedicated to the collective operations that make large-scale training possible**, and must deliver the following: *Extreme bandwidth density* to the tune of terabits per second per node; *Microsecond-level latency and low jitter* because every iteration is gated by the slowest GPU; *Near-zero packet loss* – to avoid retransmissions that stall hundreds of GPUs; and *In-order, congestion-aware delivery*

Historically, InfiniBand (IB) has been the primary choice for back-end GPU networks because of its ultra-low latency and high bandwidth. However, IB solutions significantly increase both capital and operational expenses due to higher hardware costs (50–100% more than Ethernet-based networks) and complexity in maintaining separate front-end and back-end infrastructures. Additionally, reliance on IB often locks customers into a single vendor, reducing flexibility.

Modern AI networks face a surprising weak spot: Network reliability. A single 10 k-GPU pod can expose hundreds of thousands of 400–800 G fiber links - links that are fragile! A speck of dust, a few extra degrees of heat or a minor misalignment can degrade the optical power just enough to trigger throttling or "flapping". Field data show AI clusters suffer up to 10X more flaps than web-service fabrics, slashing effective GPU utilization.

AI's Unique Fabric Requirements Translate Into Some Major Challenges

Modern AI networks are engineered for lossless, terabit-scale bandwidth and micro-second latency, yet those same requirements expose three stubborn gaps that routinely derail production workloads.

- **Visibility gap**: Teams lack real-time visibility into network conditions such as link availability, congestion and contention on links, packet-level and message-level latency, and bandwidth utilization. As a result, teams are unable to detect or diagnose issues often until application performance suffers.
- **Resiliency gap**. AI clusters suffer network failures and flaps 10 times more often than web-service fabrics, many of which can force jobs to be restarted from the last checkpoint, causing cluster underutilization and delayed job completion times.
- **Performance gap**. AI workloads exhibit bursty traffic patterns, causing flow collisions during NCCL operations (e.g., all-reduce or all-to-all operations) or during simultaneous inference bursts. This increases latency, reduces throughput and ultimately impacts job completion times and token streaming performance.

Together, these challenges lead to an AI infrastructure efficiency gap. Low goodput and high p99 collective latency translate directly into poor Model FLOPs Utilization (MFU). Limited infrastructure visibility - particularly at the network layer - compounds the challenge by making root cause analysis harder and extending mean time to recovery. The result is thousands of GPUs idling, dragging ROI, power utilization and model iteration down.

The future: Software-Driven Fabric ("SDF")

AI infrastructure is poised to replay the industry's long-running shift from vertically integrated mainframes to disaggregated, best-of-breed systems. NVIDIA's soup-to-nuts stack - Hopper GPUs, NVLink/NVSwitch fabrics, Mellanox InfiniBand, Spectrum Ethernet switches, and CUDA - delivers stunning performance but binds customers to a single vendor. As model sizes soar and budgets tighten, we are already seeing the emergence of custom accelerators from AMD, Google, AWS, and Microsoft at the silicon layer, and vendor-neutral runtimes such as PyTorch + XLA, MLIR, and Triton to abstract the software tier.

Ethernet is positioned to gain ground as the networking fabric for AI clusters. Historically, Ethernet's expansive ecosystem and decreasing costs have outpaced specialized networking technologies, such as ATM in carrier networks and Fibre Channel in storage. With advancements like 800 G/1.6 T lanes, lossless RoCE (RDMA over Converged Ethernet), and improved congestion management, high-speed Ethernet is already complementing InfiniBand inside many GPU clusters. As training and inference workloads coexist on the same cluster, the opportunity to drive efficiency by converging front-end and back-end networks will become more compelling. ROCE is already being used as the transport for AI optimized storage systems that used to mainly reside on front-end networks. As volumes rise, Ethernet will swallow ever larger tiers of AI fabrics ultimately outpacing InfiniBand over time.

Crucially, Ethernet brings with it the software-defined playbook that unlocked agility in cloud data centers. Separating control and data planes allowed SDN to carve a flat network into job-level virtual networks, enforce micro-segmented security, or reroute traffic around hot spots in seconds. **That same software-driven dynamism is required for AI workloads - code drives network behavior in a hardware agnostic manner**.

Clockwork believes that a Software-Driven Fabric, with the following core principles, is a key building block of modular, multi-vendor AI platforms:

1. **Software-Driven**: Hardware and vendor agnostic - Any GPU, Any NIC, Any Switch, Any transport (IB, ROCEv2, Ethernet)
2. **Granular telemetry**: Pervasive, real-time, sub-microsecond-accurate network telemetry
3. **Closed-loop autonomy**: Failover, congestion control, rerouting, etc. driven by real-time telemetry
4. **Designed for diversity and scale**: Scale to 100s of thousands of accelerators and links, spanning GPU providers, Ethernet or Infiniband transport, switch providers and NIC providers
5. **Application-aware**: Software-driven segmentation/QOS ultimately facilitates 1 physical network

To realize this vision of a truly software-driven, modular, and multi-vendor AI infrastructure fabric, foundational innovations are essential. Clockwork's Software-Driven Fabric achieves these ambitious goals through two fundamental technological pillars - Global Clock-Sync and Dynamic Traffic Control - ensuring that software can effectively orchestrate network behavior at unprecedented scale and precision.

Clockwork's SDF Platform Foundations: Global Clock-Sync

Global clock-sync creates a unified, sub-microsecond-accurate timeline for high resolution network telemetry

Clockwork's Software-Driven Fabric (SDF) brings nanosecond-grade visibility to AI infrastructure, delivering three integrated views: Fleet Audit, Fleet Monitoring, and Workload Monitoring.

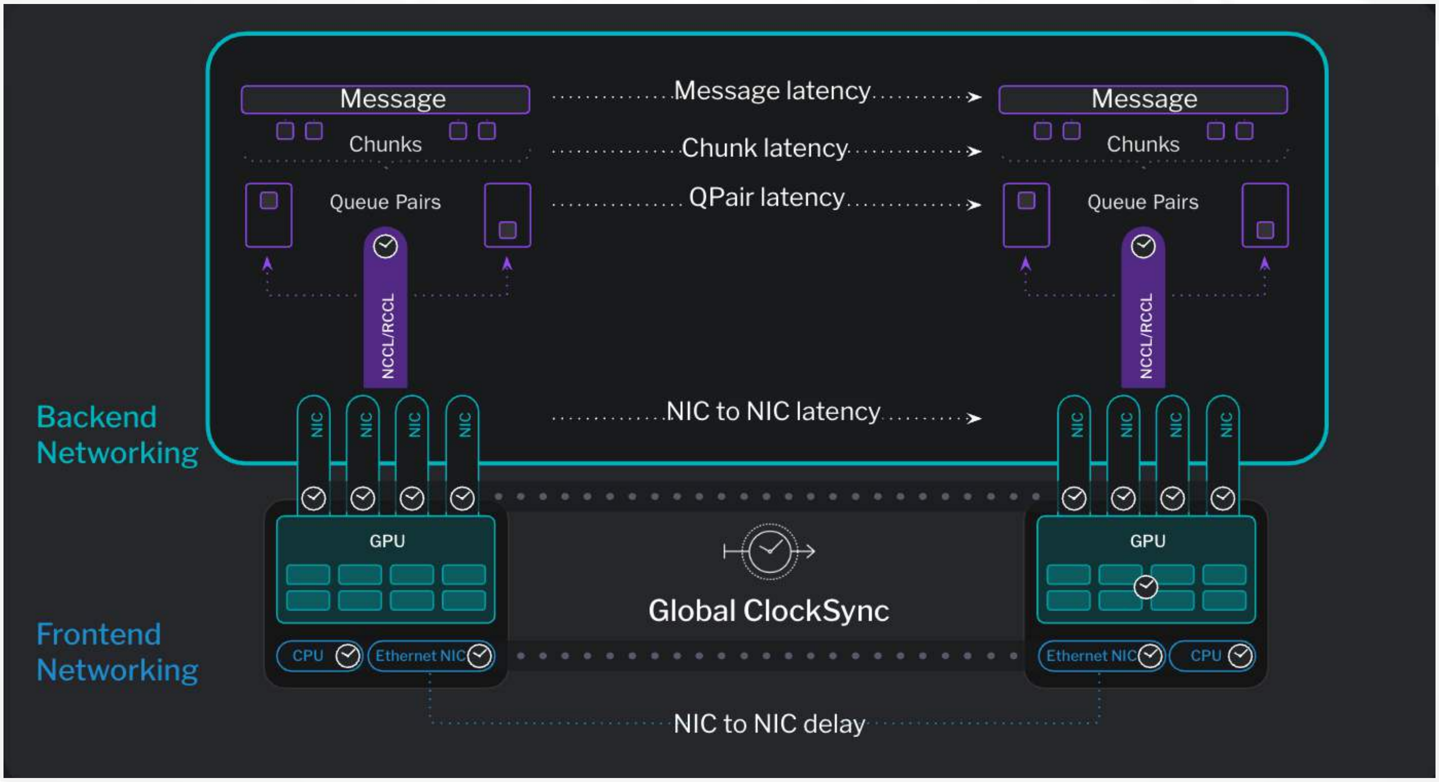
The platform builds on two pieces of proprietary R&D: (1) **software-based global clock-sync**, which aligns every host, switch, and SmartNIC to a common timeline with sub-microsecond accuracy; and (2) **edge-based queue-inference algorithms** that extract path and queue dynamics directly from packet probes. Tens of thousands of nodes—within one data center or across regions—can therefore operate against a unified nanoseconds-accurate timeline!

With this shared timeline, the same lightweight probe mesh produces an ultra-dense telemetry fabric. Operators see their network from two angles at once: *provisioning health* (Are the rails wired and firmware correct?) and *runtime efficiency* (Which jobs or links are becoming outliers?). Alerts fire on the first sign of trouble—link, node, or workload—before users notice a slowdown.

- **Fleet Audit –** validates every node and link before workloads run: software versions, firmware/driver status, PCIe and NVLink health, SmartNIC reachability, end-to-end network connectivity checks, topology assessment and NCCL path throughput checks.
- **Fleet Monitoring –** Continuous NIC-to-NIC one-way-delay measurement streams drive real-time heat-maps of path latency, hop-level congestion alarms, link-utilisation graphs and automatic "slow-sender / slow-receiver" alerts - scalable to millions of Smart-NICs without mirroring any user traffic.
- **Workload Monitoring –** A lightweight NCCL/RCCL plug-in tags production packets so the same nanosecond timeline is applied inside jobs: queue-pair and chunk-level one-way delays; per-collective throughput; ECN and instantaneous GPU-to-GPU flow composition.

Mis-wired rails, GPUDirect-RDMA mis-configurations or uneven spine contention show up immediately as outlier flows, letting infrastructure engineers and data science teams to drill from cluster map → rack → queue-pair → offending job in a few clicks.

Clockwork FleetIQ Platform Foundation: Global ClockSync Delivers Insane Visibility



Clockwork's SDF Platform Foundations: Dynamic Traffic Control

Dynamic Traffic Control enables software-driven flow and path management for network failover, congestion and contention management, QoS and segmentation

Clockwork's **Dynamic Traffic Control (DTC)** manages network traffic in real-time purely through software. DTC automatically optimizes paths, mitigates congestion, and quickly adapts to network disruptions without manual intervention. Because DTC is software-based, it works seamlessly across diverse hardware setups - multi-vendor Ethernet networks, Infiniband networks or ROCE deployments, supporting compute clusters powered by NVIDIA, AMD, and other accelerators.

DTC can orchestrate traffic at multiple layers enabling granular control across the entire communication stack - dynamically pacing or delaying individual packets based on real-time network conditions. Through integrations like NCCL plugins, it can steer queue pairs, reroute collective operations, or optimize message distribution across the fabric. Since the control logic runs in software, new traffic strategies and heuristics can be introduced without any changes to underlying switch hardware or firmware.

Some of the key features that DTC enables include the following:

- **Workload failover**. Network failures, particularly link flaps and NIC failures, are alarmingly common in GPU clusters, driven by high link speeds, thermal stress, and the scale of modern deployments. These failures can bring distributed AI training to a halt, as collective operations like all-reduce depend on every node participating without interruption. Clockwork addresses this challenge with real-time failover powered by deep, synchronized telemetry. Its out-of-band probe mesh continuously monitors NIC-to-NIC paths, using nanosecond-accurate one-way delay (OWD) measurements to detect even transient failures. Failures trigger fast, software-driven rerouting of affected queue pairs to healthy paths — all without human intervention.

  This "failover without failure" model ensures that jobs continue to make forward progress, even during network degradation, and rebound to full speed the moment the network heals.

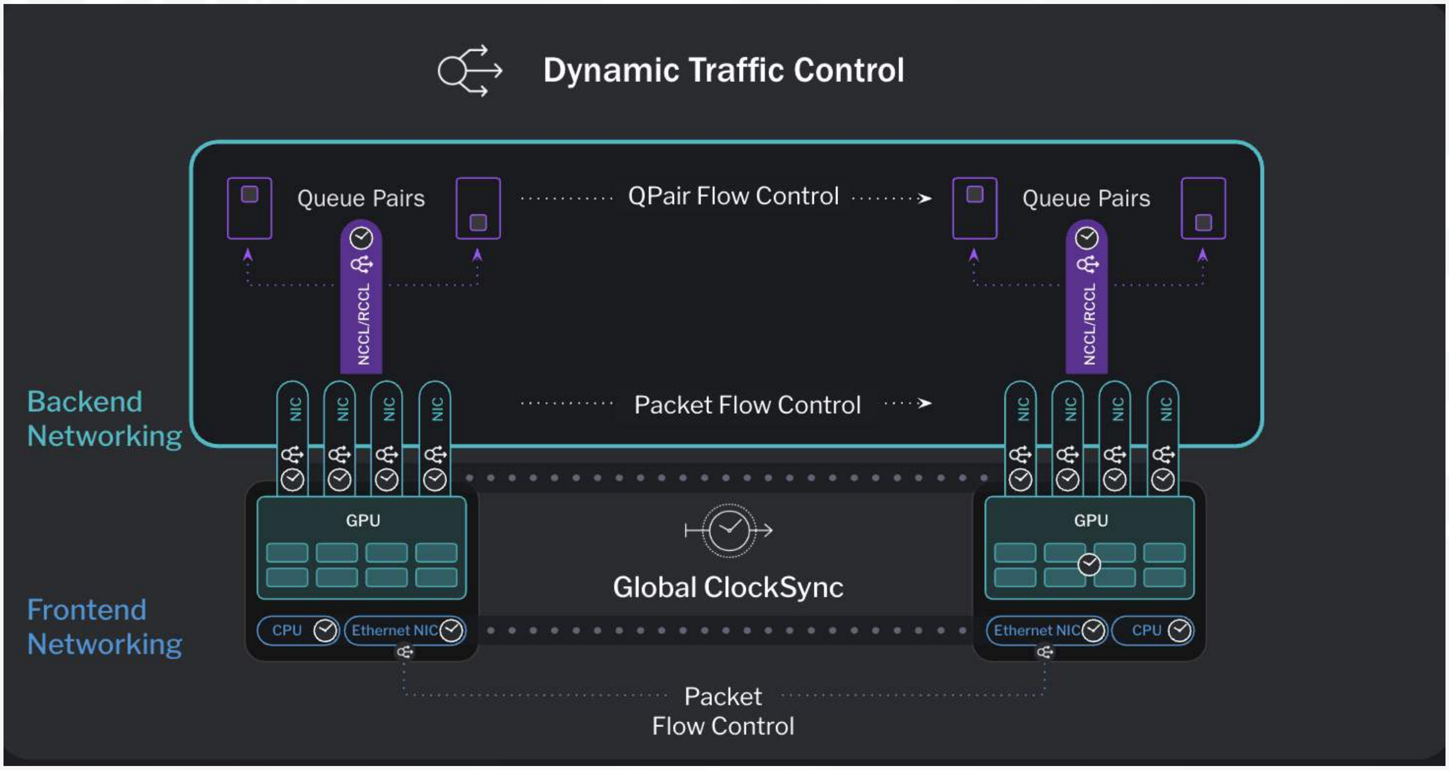
- **Workload acceleration through congestion management and contention avoidance**. Clockwork's dynamic control extends beyond failure recovery to actively accelerating fleet performance by managing congestion and avoiding flow-level contention - two key impediments to AI job efficiency at scale.

  Clockwork continuously aggregates telemetry from its probe mesh, including one-way delay (OWD) measurements and ECN (Explicit Congestion Notification) marks, to detect incipient congestion or contention on a path. In dense AI clusters, simultaneous large all-to-all jobs can unknowingly collide on the same spine links, severely degrading throughput. When delays spike disproportionately on a path, Clockwork intelligently steers a portion of the traffic from one of the colliding jobs to a less congested, underutilized route. Beyond contention avoidance, it can selectively throttle the sending rate of affected flows with an intelligent, software-driven congestion control loop layered atop the underlying fabric, thereby bounding tail latencies in the event of incast congestion.

  By combining congestion sensing with real-time, flow-aware path optimization, Clockwork accelerates the entire fleet, maximizing GPU utilization and ensuring consistent job throughput even under the most demanding AI workloads.

- **Workload QoS.** For multi-tenant or mixed-workload environments, Clockwork enforces QoS via Bandwidth Slicing - its software-defined solution to carve networks into isolated logical segments. Like cloud VLANs but without hardware dependencies, Bandwidth Slicing guarantees resources for critical workloads (e.g., shielding latency-sensitive inference from bulk training traffic), enforcing isolation and performance guarantees on commodity hardware.

Clockwork FleetIQ Platform Foundation: Dynamic Traffic Control Delivers Network Failover, Congestion Control and Load Balancing



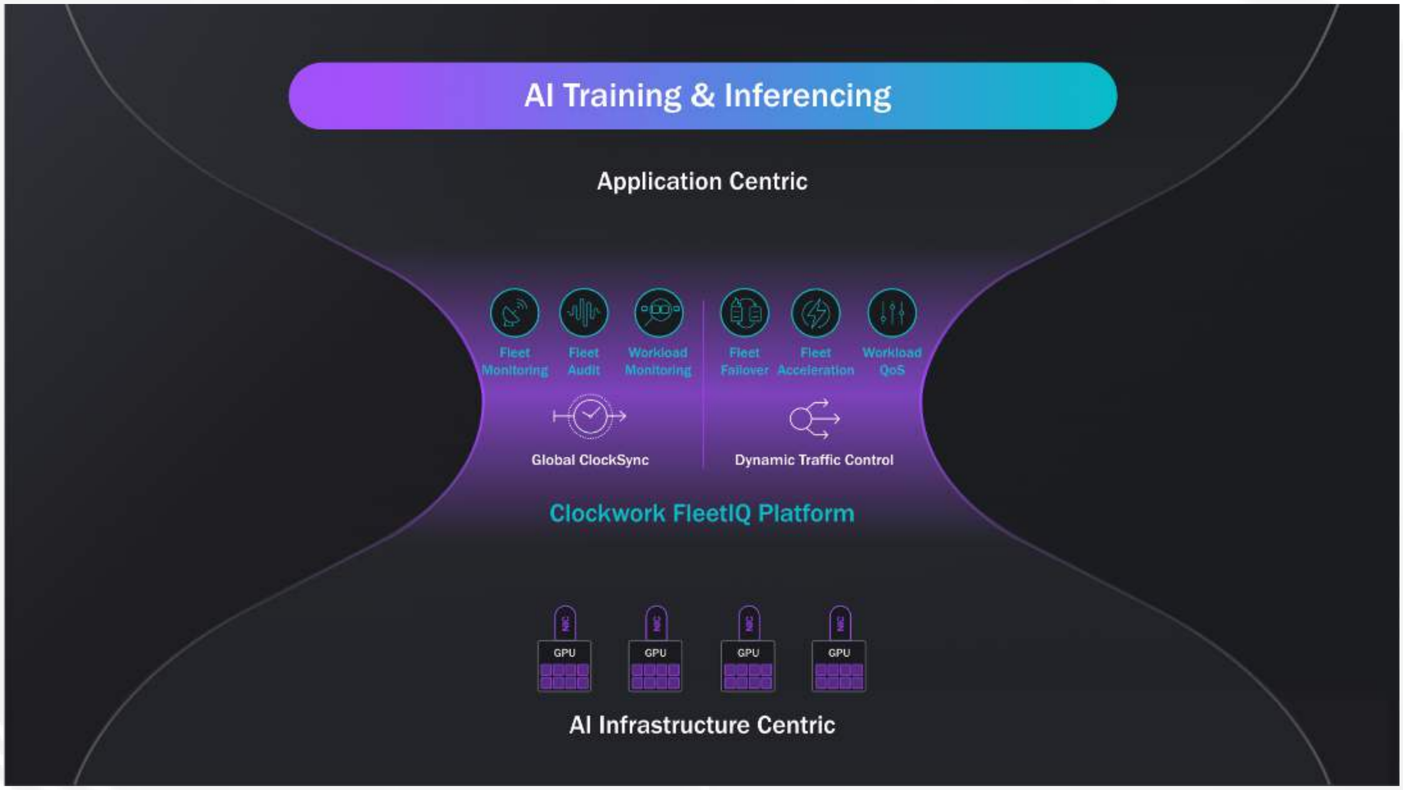
Operational and Business Benefits of Clockwork's SDF

Clockwork's Software-Driven Fabric (SDF) delivers substantial operational advantages that translate directly into improved business outcomes, particularly in terms of cost savings, performance gains, and resource efficiency.

- **Reduced Downtime and Improved Reliability**:Clockwork's Software-Driven Fabric (SDF) significantly enhances network reliability, directly reducing downtime in AI clusters. SDF continuously monitors network conditions with nanosecond precision, enabling immediate detection of link flaps, NIC failures, or other disruptions. By automatically rerouting traffic around affected areas without human intervention, SDF prevents the costly interruptions typical in large-scale AI environments. As a result, cluster utilization and operational efficiency increase substantially.

- **Enhanced GPU Utilization and Performance:** Clockwork's SDF directly improves GPU utilization, one of the most significant costs in AI infrastructures, by intelligently managing network traffic and reducing bottlenecks. Its Dynamic Traffic Control (DTC) continuously optimizes data flow, preventing network failures from causing restarts and congestion-related GPU idle times. This capability boosts Model FLOPs Utilization (MFU) meaningfully, meaning resources are utilized more effectively, and AI model training cycles are completed significantly faster.
- **Scalability and flexibility.** Clockwork's SDF architecture inherently supports scalability and adaptability, seamlessly accommodating multi-vendor hardware - multi-vendor GPUs, Ethernet or ROCE or Infiniband networks. Clockwork's bandwidth slicing technology provides effective workload isolation, guaranteeing consistent performance in mixed-workload scenarios. This flexibility ensures that infrastructure investments remain valuable and adapt to future advancements in AI platform technologies.
- **Cost Efficiency and Lower Total Cost of Ownership (TCO)**: Clockwork's SDF enhances GPU cluster utilization while significantly lowering infrastructure costs by enabling the adoption of widely available, commodity Ethernet hardware instead of specialized InfiniBand equipment. Additionally, SDF dramatically simplifies the operational complexity and reduces the manpower required to manage large-scale GPU networks. Collectively, these benefits result in substantial improvements to the total cost of ownership.

Clockwork's Vision



Over the coming years, AI clusters will expand rapidly with diverse GPUs, custom accelerators, and fast NVMe storage interconnected by commodity Ethernet. Clockwork's Software-Driven Fabric (SDF) uses software instead of proprietary hardware to deliver resilience, determinism, and superior price-performance.

Two innovations drive this approach: Global Clock-Sync creates a unified nanosecond-accurate timeline, turning the fabric into a real-time sensor grid that instantly reveals network issues. Dynamic Traffic Control leverages this telemetry to proactively manage traffic, reroute flows, and eliminate congestion hotspots without packet loss. Operating just below the application layer and above the hardware, SDF seamlessly bridges workloads and the underlying fabric - enabling future services such as micro-segmented security, continuous network snapshots, and forensic replay.

With nanosecond visibility and software-defined control, Clockwork enables faster training, safer deployments, and more efficient infrastructure use today - not in some distant future.

To request a demo, contact us at **hello@clockwork.io**